CAESAR Offshore

P O S C
Petrotechnical Open Software Corporation

POSC/CAESAR Project

Implemenation issues

Exchange and Sharing
of Oil and Gas Facilities Life Cycle Information

1 Introduction

1.1 Document Purpose

This document gives an overview of topics related to the implementation of the POSC/CAESAR Product model. The intention of this document is to bring up some of the main issues related to implementation of the POSC/CAESAR model. It is not the intention to be a userguide or a specification of implementation of the product model.

This document gives a few examples of implementation architectures. Currently no conclusions have been made regarding the best way to implement the POSC/CAESAR product model.

For further information regarding the Product model, reference is made to the project documentation Snapshot C/D.

1.2 The POSC/CAESAR Project

The vision of POSC/CAESAR is to reduce the life cycle cost and development time of oil and gas facilities by improving sharing and exchanging of information among those involved in the facilities life cycle.

Objectives of POSC/CAESAR:

- Produce agreed standards Product Model for digital descriptions of facility products
- Understand and facilitate the use of available technology for implementation of the standards
- Encourage and assist take-up of the standards and technology by the business

The Product Model will provide means to transfer information way between applications or companies contributing to the life cycle of a facility. Manual copying, interpretation or rewriting of data into different application will not be needed. Data are only entered once, and are then reused throughout all stages of the facility life cycle.

1.3 Document Readership

This document is intended for the following categories of readers:

- Software application designers and programmers
- Information and data management specialists
- System integration specialists
- System providers for design and detailed engineering activities

1.4 Document Contents

This document will give a short resume of relevant technology and some ongoing projects which aims at fulfilling the POSC/CAESAR's vision of data exchange and sharing

- Storage of data
- Access to data (read/write/update)
- Export and import of data
- Exchange file format

1.5 Glossary

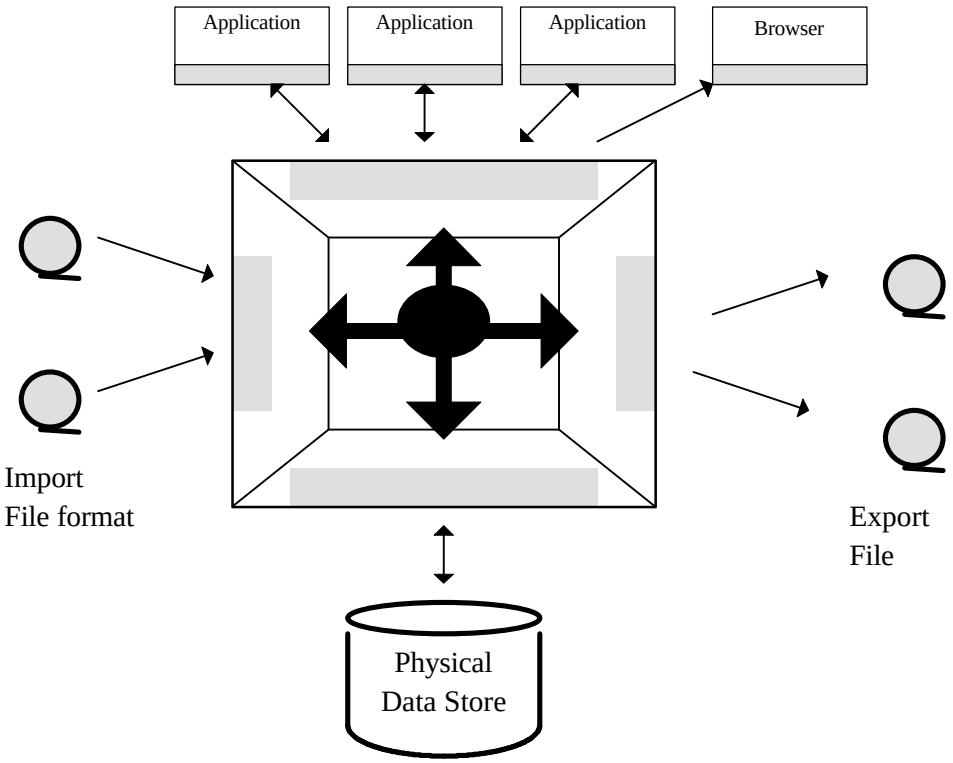
AP	Application Protocol, a part of the ISO 10303 standard that describes the use of integrated resources satisfying the scope and information requirements from a specific application context.
AP221	Application Protocol 221. Application Protocol for functional data and schematic representation for process plants
API	Application Protocol Interface
DAE	Data Access and Exchange. API defined by POSC
DAEF	
DDL	Data Definition Language.
EPICENTRE	A model for exploration and production data produced by POSC.
EPISTLE	European Process Industries STEP Liaison Executive
EXPRESS	A formal data specification language specified in ISO 10303-11.
EXPRESS-M	Graphical representation of an EXPRESS data model.
PART-21 P-21	ISO 10303 - 21. Definition of a physical file format for exchange of product data.
PEF	POSC Exchange file
PIPPIN	Pilot Implementation of a Process Plant Information warehouse. ESPRIT project.
POSC	Petrotechnical Open Software Consortium
RDBMS	Relational Database Management System
SDAI	Standard Data Access Interface. API specified in ISO 10303-22.
SMDTI	Single Model Data Transfer Initiative. Implementation of POSC technology for exchange and integration of data.
SQL	Structured Query Language
STEP	Standard for Exchange of Product data
UVI	User Visible Identifier

2 POSC/CAESAR IMPLEMENTATION

2.1 Overview

Implementation of the POSC/CAESAR data model and reference data will depend on software, databases and interface descriptions. Below are elements of a possible architecture discussed. This architecture depends on :

- Interface protocols between the different parts.
- Functions performed by different software components



Indicate areas which require standardisation to enable exchange and sharing.

An actual implementation of these ideas into software is not within POSC/CAESAR's scope, but there exists several initiatives that may cover this need and some of them will be discussed in the following parts of this document.

2.2 The POSC/CAEAR product model and reference data (class) library.

The POSC/CAEAR product model has an Express formulation. The product model keeps the reference data model as an integrated part.. The model may serve two different purposes

- Reference model for exchange of data
- Conceptual model for a system implementation

The latter could be as a data warehouse, as a central repository for discipline applications, etc.
An implementation of the POSC/CAESAR model depends on a set of reference data. Today's reference data library has its greatest coverage within material so extensions for facility, information, activity reference data is expected.

An implementation will also need other requirements to be formulated, for example conventions about interpretation of the model, constraints, security, identifiers, etc.

If the model is used as a reference model for data exchange, these requirements may not be stated.

The reference data library has structures that may prove not very efficient when used within an operational data store for an offshore installation. One aspect is the hierarchy within the reference data, which will require traversing when used. One simplification would be to process the intended inheritance, so that all reference descriptions are complete and no checking for inheritance is needed by the ware house. Other strategies should also be examined if a full structure is not necessary.

2.3 Data store

In an actual implementation the model will be used as a conceptual model, and additional requirements either added to the database or included in the interface or application layer.

The design of the data store will have to meet a set of requirements,

- Performance
- Security
- Consistency and integrity
- Change requirements, versioning

Design of the database may choose to rely on existing architecture or standards. If so, interface protocols will be dependant on a mapping between the conceptual and the internal schema. This technology is evolving, but for the time being only POSC technology has these capabilities.

Performance is concerned with extract of data import, read and write. A data store designed for data interchange will have modest requirements for response time, while a data store as a repository for discipline applications will have rather strict requirement. For the time being no known design seems to fulfil such requirements, but new possibilities and technology develop all the time. There are special hopes for a design that includes an object oriented access layer.

Security, consistency and integrity requirements will have to be developed during the implementation.

A significant problem will occur with data in databases if the meaning intended by the writer is inconsistent with the understanding developed by the reader. This is also relevant for Import and Export operations.

The mere size of the database will also have to be of concern.

Another concern is the ability to handle changes with the model and the reference data library. The model and the reference data library have not reached version 1 status yet, so updates will no doubt arrive.

2.4 API LAYER

The API (Application Protocol Interface) layer may be thought of as a layer between the user application and the data base storage management. It allows interface to data seen as through the conceptual model. This API could in some cases be omitted. In such cases the end applications will access the data store directly. For POSC/CAESAR this will require the existence of mapping between the P/C model and the data base lay out. It will require special gifts to write SQL queries against the data base. So for tasks that go beyond the most trivial direct SQL calls will not be viable.

Ongoing developments pursue different strategies

- User applications with direct access to the data store.
  These applications will benefit from less overhead, but may be dependant upon a correct understanding of a conceptual schema and the completeness and correctness of a such schema. Also redesign of the data store may have a serious impact.
- A set of applications between user application and the data store.
  The user applications will depend on the conceptual model and the API layer that provide access t to the data storage. This will require mapping between the conceptual and the internal schema, security handing, integrity management and several other requirement(see the POSC DAE specification).
- Business object
  To facilitate the access to the data store there is ongoing work to clarify use of object oriented approaches. Business object is a such development. Until now no authoritative definitions exist, but one choice could be to develop objects that encapsulate data and methods related to specific equipment; e.g. a pump. Such an object could benefit from the reference data library.
  The main reasons for such development would be to facilitate access to data, to remove doubts about the correct interpretation of the model

2.5 Export

Export of the data will depend upon on

- Selection possibilities
- Documenting the export
- Format

The selection may be restricted by different set of criteria as

- Context (electrical or mechanical descriptions of equipment, data to create a specific drawing,...)
- Possibility to limit which items to be exported
- Updates since a given time, versions

2.6 Import

The requirements for the import file is related to the requirements for the export with some additions

- documentation of content
  Data description, file lay out, origin
- Identification (UVI)
- complete/partial model
- new, update, additions
- conformance and reconciliation
- merging of data
- manipulation of incoming data

Previously to loading into the data warehouse, data should checked for

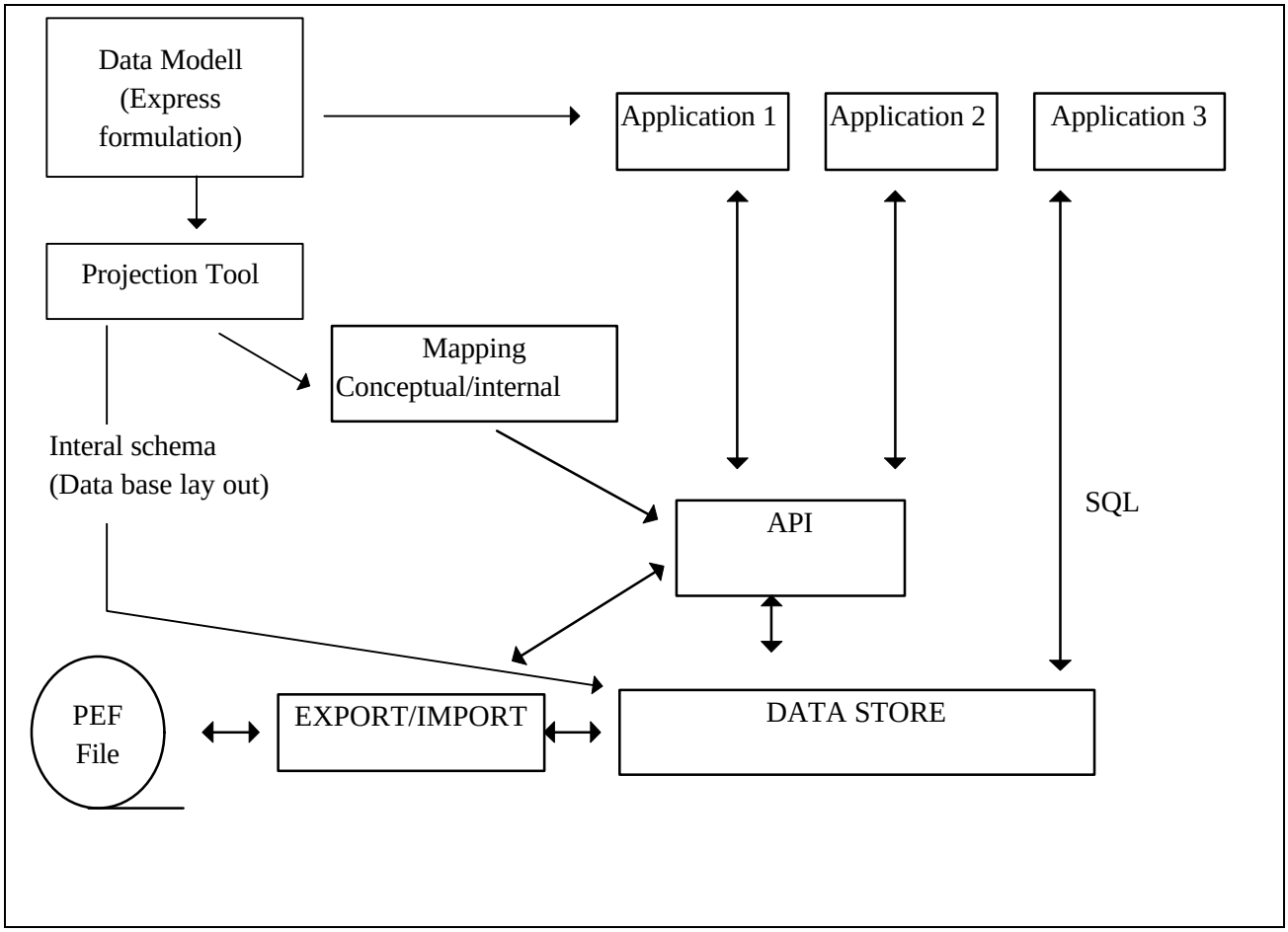
- Consistency against the exporting schema,
- Consistency against the POSC/CAESAR schema,
- Consistency the reference data library,
- Consistency integrity requirements
- Other potential constraints that may be applied by the data warehouse

3 POSC SIP

3.1 Overview

POSC has developed system architecture and issued a specification POSC **SIP** (Software Integration Platform), which consist of a set of interface specifications and requirements for the software environment. The SIP standards consist of the following specifications:

- **EPICENTRE Logical Data model**
- **DAE API specification for the logical model**
- **PEF; Posc Exchange Format**
- **Basic computer standard**
- **User Interface Style Guide**
- **Inter-Application Communication**



Overview of POSC Architecture

The SIP environment includes data, computer hardware and software. This enables the portability and interoperability of applications and data.

POSC has developed a sample implementation of this architecture and commercial development will soon be available.

The SIP specification was designed to make it possible to utilise the EPICENTRE model. The only assumption for the model was that it has an EXPRESS formulation. Thus it is possible to use the SIP architecture and general software developed for SIP within a POSC/CAESAR implementation.

3.2 EPICENTRE model

This model is an extensive entity relationship model covering most aspects of subsurface items related to oil exploration and production. The model is documented by Express.
There is app. 1500 entities within the model. A number of different basic data types has been defined based on Express primitives.
The model also contains parts that are related to the administration of data. Typically export and import of data within the PEF concept requires additional data.

The model contains explicit descriptions in contrast to POSC/CAESAR model that is a generic model and depends on a reference data library to hold real world data.

3.3 Data store

The idea is that the programmer should be able to access the data through a conceptual[1] model knowing nothing of the internal schema. Also the actual distribution of data on different computers should be hidden for the application programmer

The mapping between the internal and the conceptual model has to be made available to the API software. A special software tool called 'Projection tool' has been made to create an internal schema based on the conceptual. The mapping between these two schema is also created according the he mapping specification. This projection tool requires an Express schema and configuration data as input. This means that any schema written in Express could make use of Posc architecture and utilities.

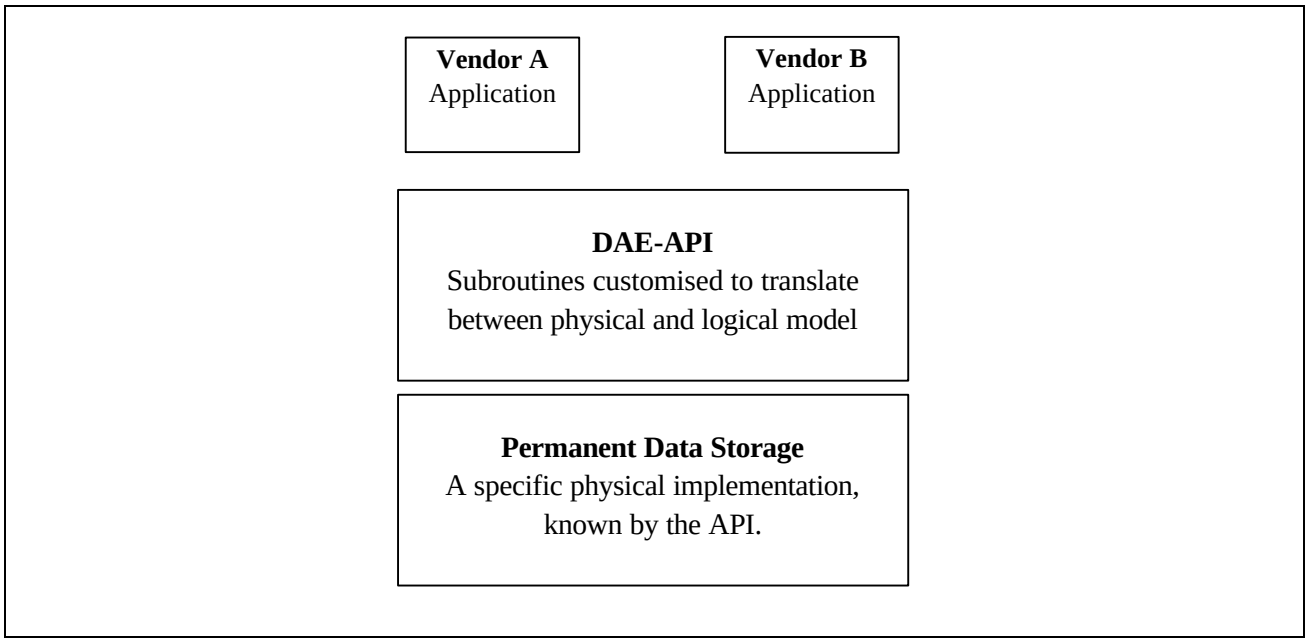
The protocol and the projection tool put restrictions on possible internal schema. The projection technique is basically a denormalisation with no denormalisation as one extreme and a 'One table' design as the other extreme. So far it has been a problem to match the existing technology, the model and the projection tool to get a satisfactory performance. Related take-up projects work with this problem.

[1] The EXPRESS definition of Epicentre is reffered to in the POSC documentation as a logical data model. POSC/CAESAR use the term conceptual data model for the EXPRESS model produced by POSC/CAESAR, but the intent is that these two terms should be regarded as having the same meaning within this document, and so from here on, the term conceptual will be used when reffereing the Epicentre model.

3.4 DAE (POSC API)

The purpose of the Date Access and Exchange application programming interface; DAE, is to provide a standard means of accessing data as specified by the data model, and managed within a POSC Data Store. The APIs define a set of calls to provide application software for accessing, creating, querying, updating and deleting data. It allows a data sharing environment in contrast to the ISO STEP that focuses on exchange of data.

Due to the complexity of model, both POSC and other initiatives in EUROPE are working on the concept of business objects. One implementation of such object would be objects that take care of their own data and methods. These could remove the need for the use of 'Foot prints' now widely used by POSC to provide a unified handling and understanding of data.



The DAE API is based upon the ISO/STEP SDAI specification, described in chapter 4.2.

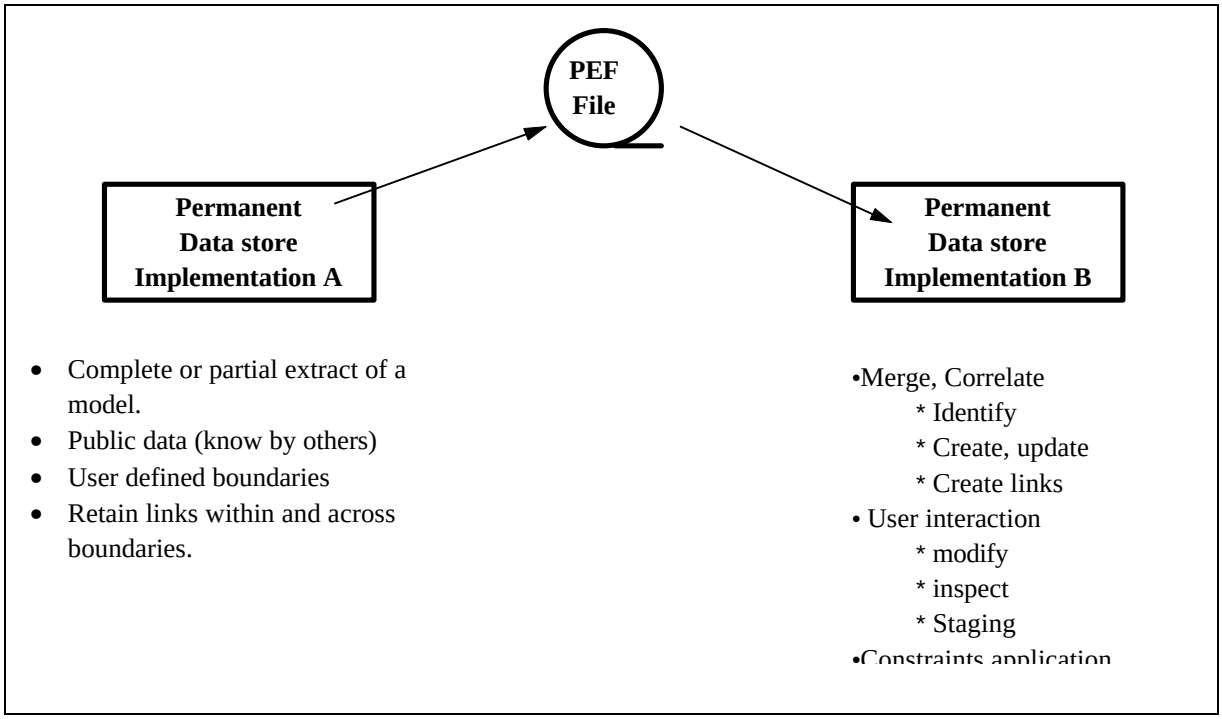
3.5 PEF (Exchange)

Data exchange is provided by the PEF (Posc Exchange File) specification. This specification defines a standard for exchange of data using a physical file as media. This specification defines the file format, and functions for extract and import of a file.

The POSC Exchange file format will make it possible to move data from one type of computer or data Base Management System to another. The PEF concept includes following options or features:

- Archiving data for future use
- - Transfer data to other sites within company
- - Transfer data from external sources.

- No dependency of storage model
- Exchange profile
- Definition of select and merge included
- Export and import history included
- Data set description
- Data



4 ISO/STEP

The ISO/STEP focus mainly on exchange of information. The relevant standard is ISO standard 10303, called STEP(Standard for the Exchange of Product Model Data)

The STEP standard is divided into several separate Parts, where each Part is more or less a 'stand-alone-standard' and approved separately.

This section will give a brief overview of the relevant ISO STEP Parts for POSC/CAESAR Project.

4.1 EXPRESS

The ISO 10303 Part 11, The Express Language Reference Manual, is a specification of the EXPRESS data specification language. This language is used to describe all data models within the ISO/STEP standard. EXPRESS is also used for the specification of the POSC and POSC/CAESAR product models. POSC use a subset of the EXPRESS language, and have defined a few additional data types.

EXPRESS-I is a language to define instances of a data model defined in EXPRESS. The purpose of EXPRESS-I is mainly to be used during test loading of data, and not an alternative for the Part-21 physical file exchange standard.

EXPRESS-X, a procedural language to define mapping between to data models defined in EXPRESS. This standard is not finalised yet.

4.2 Application Protocols (Conceptual models)

ISO/STEP focus on exchange of information between systems and neutral storage of information. In order to achieve non ambiguity the ISO community defines Application Protocols (AP) for an engineering application area. These application protocols consists of two parts; ARM and AIM.

The ARM (Application Resource Model) model is formulated in EXPRESS that describes the application area in terms of an expert in the application area.

The AIM (Application Integrated Model) is how the ARM model should be expressed by a system design expert. Great effort is taken to harmonise different application areas that use the same concepts or handle the same real world entity from a different view angle.

Each AP is defined as a separate Part of the ISO/STEP standard.
The relevant AP's for POSC/CAESAR project are:

- AP221 Functional data and their schematic representation for process plants. (PFD, P&ID)
- AP227 Plant spatial configuration (3D CAD)

- AP231 Process Engineering Data. (Process simulation)

The scope of POSC/CAESAR is overlapping the scope of these AP's.

The AP221 is based upon the same data modelling principles as POSC/CAESAR, the Epistle Generic Entity Framework. An agreement with the Epistle organisation regarding development of a common 'Reference data library' have been signed. However the libraries may differ as the scope of AP221 and POSC/CAESAR are not total overlapping.

The Application Protocols may be used as a basis for population of a POSC/CAESAR data storage from external systems.

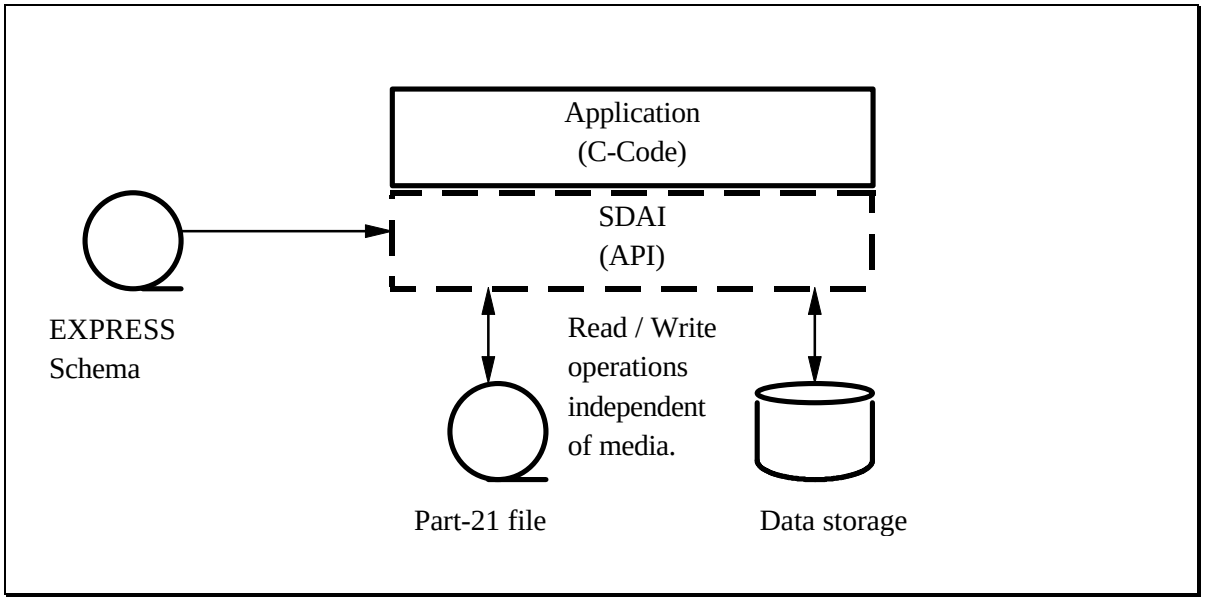
4.3 SDAI (API)

The ISO 10303 Part 22; STEP Data Access Interface, is a specification for a data access interface independent of physical storage. It is a specification of functions that enable access to data/instances where the data structure are defined in a corresponding EXPRESS schema/data model.

The ISO-10303-22 have a companion document, ISO-10303-24, which is a specification of the SDAI - C language late binding.

A physical implementation of the ISO-10303-22 could be a library of C-functions. Applications may use these functions to access a physical Part-21 file for read and/or write.

The specification of the SDAI does not include:
- Multi-user access to SDAI sessions,
- Repositories, or data by multiple applications,
- Transactions
- User definitions.



4.4 Part-21 (Exchange)

The objective of the ISO 10303-21 standard (Part-21), is to provide an exchange structure format based upon a clear text encoding for product data where the data specification is described in EXPRESS, ref. chapter 4.1.

A Part-21 file contains two main sections.

- The first contains a data model documented with an EXPRESS schema.
- The second has the data in ASCII text according to the EXPRESS schema.

The use of Part-21 format files is not limited to product model data or ISO/STEP Application Protocols. So it may be used for exchange of any information where the corresponding data model is available in EXPRESS schema format. It is straight forward to generate an Express-schema from any DDL defining a relational schema.

The Part-21 standard is mainly a description of the exchange file format, and does not give any description of how to manage exchange operations. The POSC PEF exchange standard is an example of the latter.

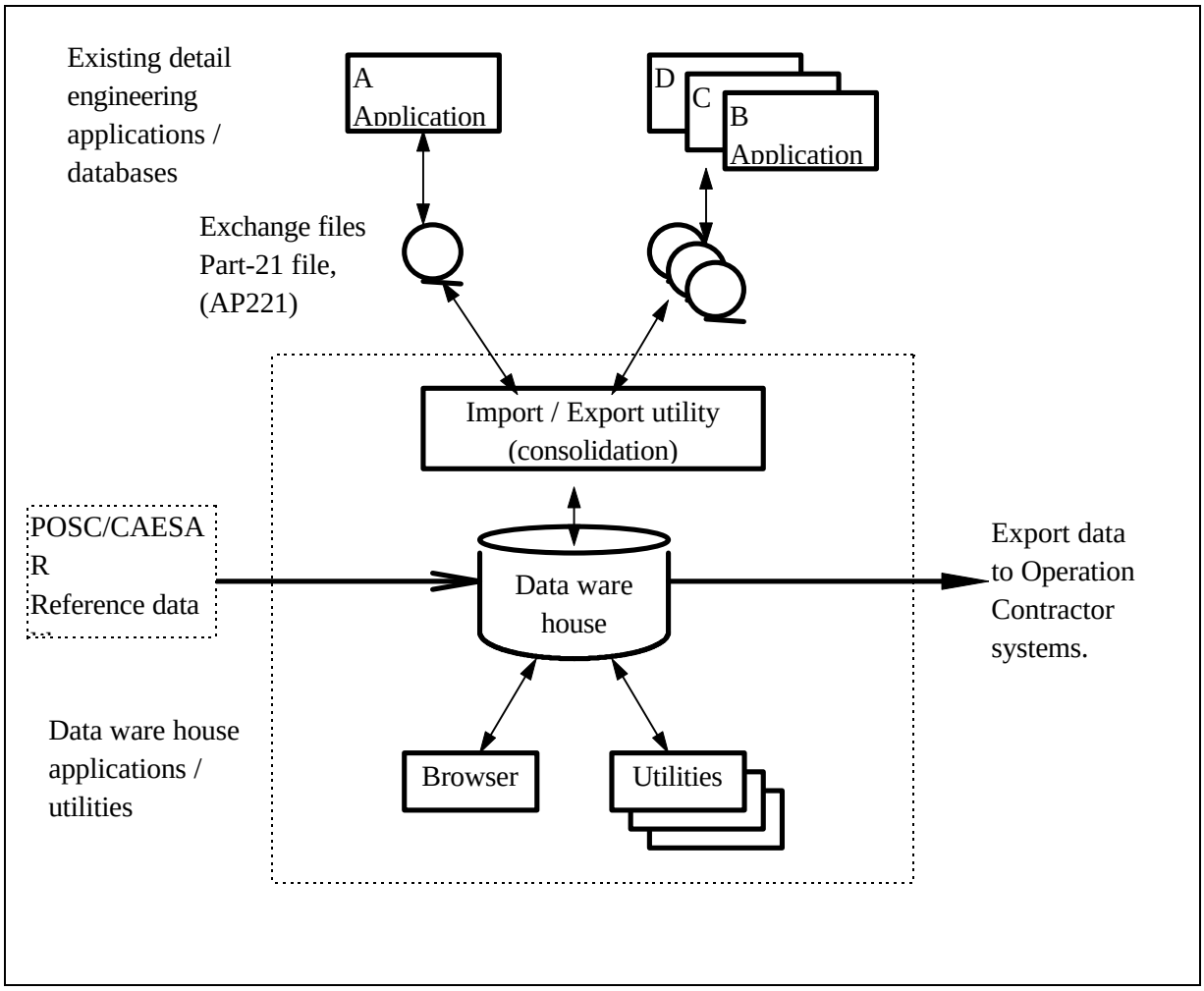
5 Related implementation initiatives

5.1 ETAP Data warehouse project

ETAP; Eastern Trough Area Project, is a field development project on the English shelf, operated by BP. The projects intend to establish a data warehouse, which is a repository of information generated by the suite of electronic data bases and design applications used in the engineering phase of the project. This data warehouse should comprise all relevant information for handover to the operation phase.

Intelligent Computer Solutions (ICS) won the contract to make an implementation of such 'data warehouse'. The first deliverable from ICS will be ready for implementation Q1-97.

The data warehouse will be based upon the AP221 data model, and the POSC/CAESAR reference data (reference data library).



5.2 PIPPIN Project

The PIPPIN Project is a EU funded as part of the ESPRIT IV program.

Budget : 17 man years
Schedule : December 1997
Participants : BP, Brown & Root, Shell, EuroSTEP, ICS, ICI, Framatome, Enator
Deliverable : Implementation architecture specification
: Data warehouse, API, Utility software

Business Objective:

- Identify business benefits from improved utilisation and sharing of data
- Stimulate market for STEP compliant Software.
- Stimulate the use of STEP in major engineering and management of assets information through life cycle.

Project Objectives:

- Design and implement an open architecture including a data warehouse and API's to a number of engineering applications.
- Build a warehouse management system to check in and check out of shared information by co-operating applications to and from the DWH
- Use delivered SW on two engineering projects (ETAP, SAK)
- Evaluate the usefulness of STEP standards (AP221, AP227 and AP231) and delivered software.
- Contribute to setting the direction and priority for further efforts to improve information management.

5.3 VÅV Project

The VÅV Project is a co-operation between the Norwegian offshore development projects;

- Varg operated by Saga Petroleum
- Åsgard operated by Statoil
- Visund operated by Norsk Hydro.

with the objective to develop a data warehouse solution, based upon the POSC/CAESAR product model.

The VÅV Project has a scope to develop a data warehouse based on POSC/CAESAR

5.4 Epistle Implementation forum

EPISTLE; The European Process Industry STEP Technical Liaison Executive, is an informal group of representatives from the Process Industry who meet to define and analyse data models for the Process Industry.

EPISTLE is responsible for the development of the AP221 standard and has three main groups

- Data modelling group
  Responsible for the development of the AP221 data model, maintaining the Epistle Framework, and the Epistle Core Model.
- Reference data Library group
  Responsible for the development of the reference data library associated with AP221.
- Implementation group.
  The Implementation group was formed due to a common interest in finding implementation related to the AP221 standard, and also EPISTLE like data models.

6 Data conversion

Having data within one model and wanting to transport it to another model involves

- Mapping between two different schema
- Transformation of data between two data bases instances.

Mapping involves handling with following issues:

- Defining what target and source schema
  (DDL, EXPRESS, Data model language,...)
- Handling of concepts that may exist within one model and not the other (classes, facility,...)
- Non conforming classification schema and philosophy
- Ways of documenting a mapping

Actual transformation of data is transformation of output files to input files

- Mapping schema
- Format of export and import files
- Necessary handling of source data
- Non conformance handling; Data within the source data may deviate from the conceptual schema.

6.1 Mapping

Mapping is a correspondence between concepts in two different schemas.

So the first thing will be to find which schema to map.

POSC/CAESAR has a conceptual model documented by EXPRESS and this model is well suited for mapping.

Legacy system will often not have a well-documented conceptual model and a mapping activity may be based on the internal (physical) schema. This schema may be defined by the DDL, EXPRESS or another suited language or methodology. Data stores based on ISO/STEP Application protocols or Epistle-like models will usually have a conceptual model defined by EXPRESS.

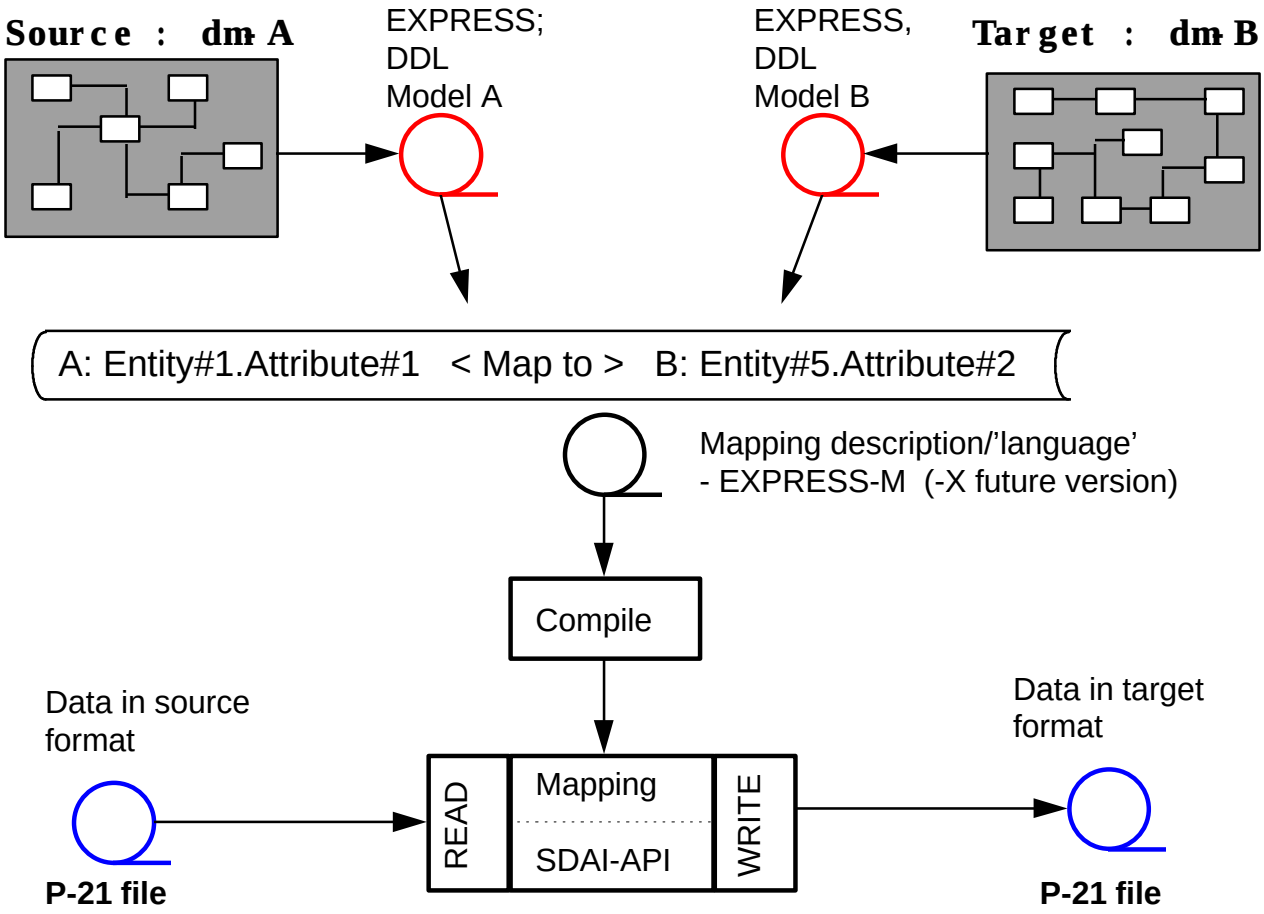
With formal descriptions of the source and the target schema it is possible to formalise the relation between the two data stores. The mapping will partly be concerned with the data in the reference data library and the explicit data elements in a non-POSC/CAESAR schema and partly with the entity and associations structure.

This documentation may use EXPRESS-M, which presuppose that the source and target schema is base on EXPRESS. Other technologies may be use the DDL if an EXPRESS formulation of the schema is not available.

Special attention must be paid to the correspondance reference data library and the subtypes within the legacy system. If there is no immediate mapping, special protocols have to be created.

Experience shows that mapping of data may be a difficult and time consuming task, especially when there is no one-to-one mapping between concepts in the two schema. The application of reference data library within POSC/CAESAR requires special considerations.

A Mapping session using the STEP technology is indicated in the figure below.



The figures show a mapping of data from system A to system B. The source and target data models (DM-A and DM-B) are described in EXPRESS. The mapping between the source DM-A and target DM-B is described using the EXPRESS-X language. The EXPRESS-X file and the EXPRESS definitions of the source and target schema are used by a 'mapping application', which read the source data (e.g. in P-21 format) and translate the data and write the data into the target format (e.g. P-21 format).

6.2 Transformation

Given a mapping between the two schema a conversion of the data may take place.

Some technologies will be able to generate conversions routines based on the source schema, target schema and the mapping. This technology my require special pre-processing to take care of the reference data library.

The conversion routines may also be written based on the mapping.

7 Conformance/ Compliance

Conformance and compliance to POSC/CAESAR are to define some equivalence classes for soft ware and data. An official statement is not available yet, but some remarks but could be given.

One part is the ability to keep data. So conformance would be that a conceptual model is compliant with POSC/CAESAR when it is possible to have a one-to-one mapping between the concepts in the two models.

To ensure interoperability there is requirements that may be fulfilled by the software involved..

Exchange files will have be conformant with the schema and the prescribed layout on the file. Special precautions will have to be taken if data from an external source have a different set of constraints compared to a P/C ware house.